The Common European Language Data Space

Online Stakeholders' Workshop

Agenda

08.50 – 09.00:   **Opening of the Webex conference room**

09.00 – 09.30:   **Welcome** and **Scene-Setting**

09.30 – 10.50:   **Tour de table/Exchanges/Q&A**

10.50 – 11.00:   **Conclusions** and **wrap-up**

Why are we here today?

A **new opportunity** for more **language data**, **tools** and **services** and for their **monetisation**.

Stakeholders

We will present…

- what the Language Data Space (LDS) is

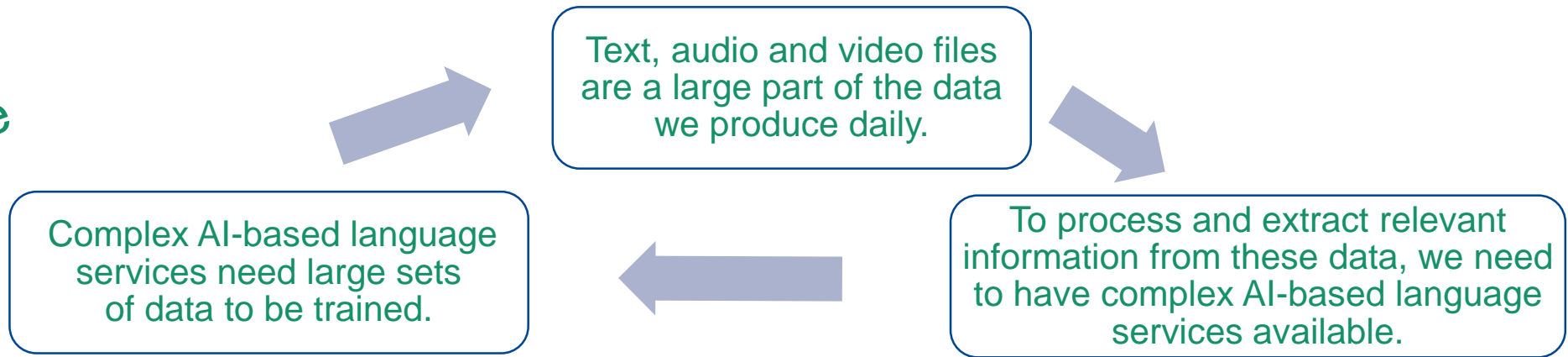- how stakeholders can benefit from it

The Common European Language Data Space

You can share…

- insights into your specific needs to be met in the LDS

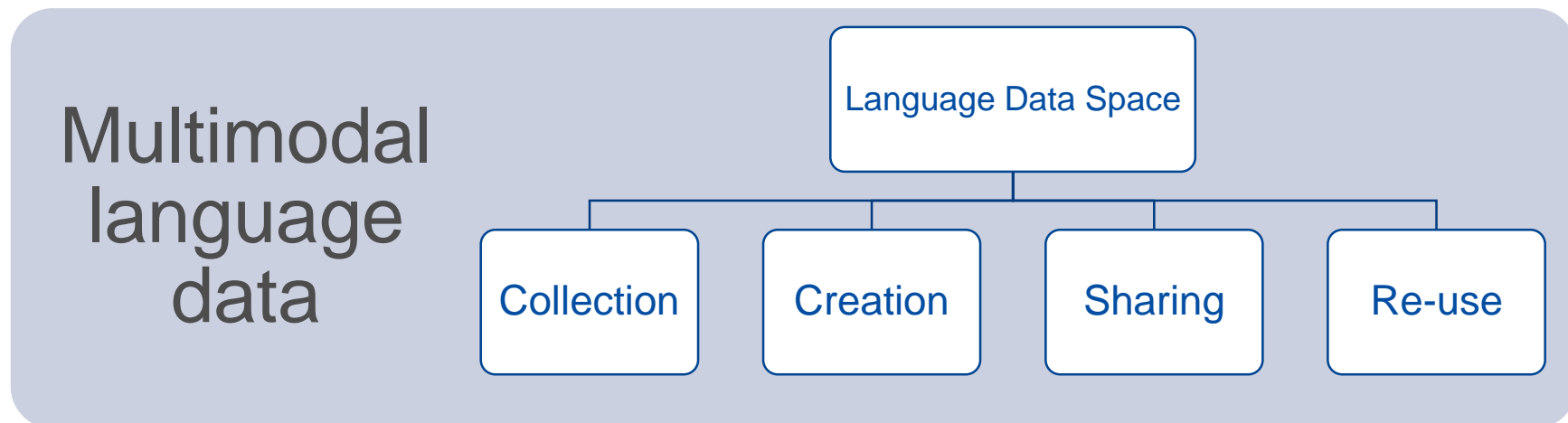- your thoughts on the LDS set-up and how to improve it

Cross-fertilisation for a mutual gain

European Commission

Language Data Space… What is it?

Increase Data Exchange

Goal

Where we are

Text, audio and video files are a large part of the data we produce daily.

Complex AI-based language services need large sets of data to be trained.

To process and extract relevant information from these data, we need to have complex AI-based language services available.

What we offer

Multimodal language data

Language Data Space

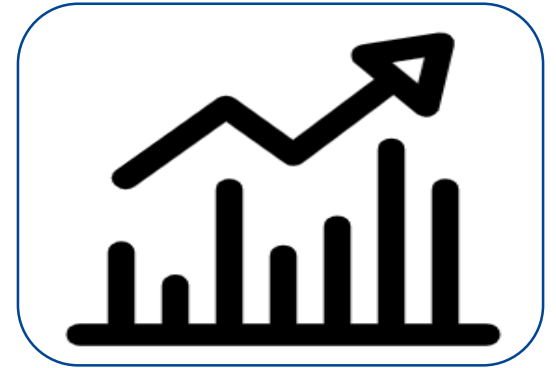Collection | Creation | Sharing | Re-use

What's in it for you
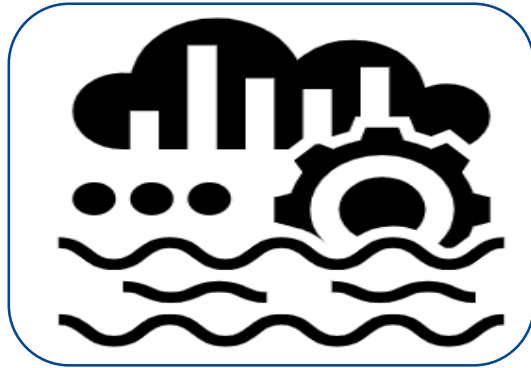
Data sharing against remuneration

Developing, selling and purchasing language-based applications, tools and models

Developing, selling and purchasing language-based services

Performing data analytics and further exploiting the market for more reach and visibility

E.g., multilingual and/or multimodal data

E.g., automatic speech transcription

E.g., automatic text summarisation

E.g., historical data

European Commission

Your involvement in the LDS initiative

Short-Term
Today's Meeting

- Presenting and exchanging ideas on the LDS to give further shape to the project

Medium-Term
June – September 2022

- Finding out more about the LDS (call for tenders publication)
- Taking part in (consortia) proposals

Long-Term
From January 2023 onwards

- Participating in the deployment of the LDS
- Using the LDS as a platform and marketplace
- Being part of the LDS community

European Commission

'Origins' of the Common European Data Spaces

Common European Data Spaces

are the centrepiece of the

European Strategy for Data

They are meant to harness the **value of data** for the benefit of the European economy and society.

Objectives

In February 2020, it announced the Commission's vision to empower businesses, public administrations and individuals to make better decisions **through the use of data**.

The creation of EU-wide common, interoperable DSs in strategic sectors is to **overcome existing legal and technical barriers** to data sharing and, as such, **unleash the potential of data-driven innovation**.

→ More data shared in a privacy-compliant, trustworthy, and secure way.

→ A genuine single market for data.

Framework Programme of DSs and LDS

The [Digital Europe Programme (DIGITAL)](#) provides funding for projects in five crucial areas:

SO3 — CYBERSECURITY

SO1 — HIGH PERFORMANCE COMPUTING

SO4 — ADVANCED DIGITAL SKILLS

DIGITAL EUROPE PROGRAMME

SO5 — BEST USE OF DIGITAL TECHNOLOGIES

SO2 — CLOUD, DATA AND AI

Programme *VISION*
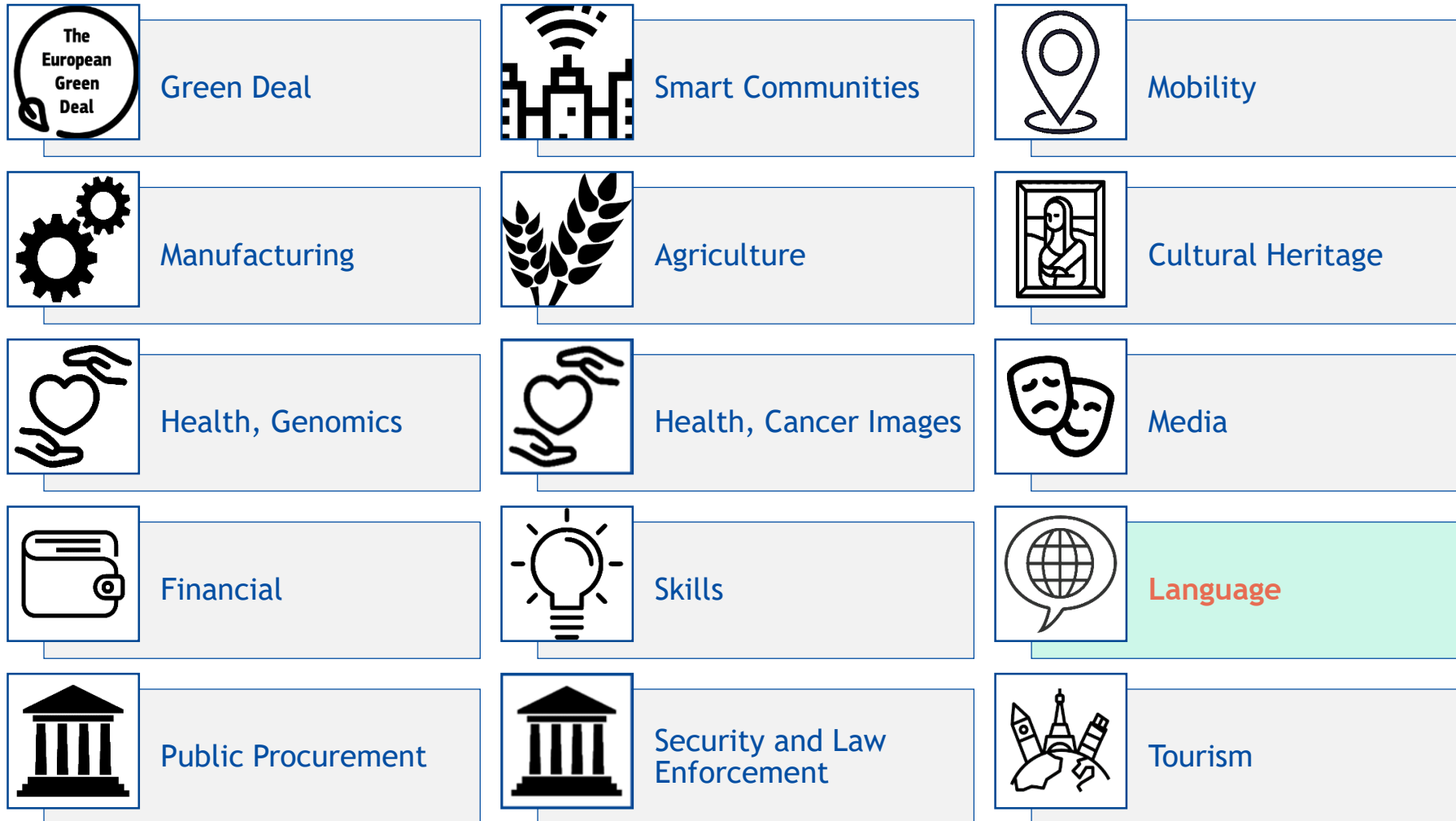Digital Decade and digital targets for 2030.

Programme *FOCUS*
Building the **strategic digital capacities** of the Union and facilitating the **wide deployment of digital technologies**.

Programme *GOAL*
Bridging the **gap between digital technology research** and **market deployment**, while supporting the EU **twin objectives** of a green transition and digital transformation.

DSs and the LDS are under SO2 'Cloud, Data and AI', whose aim is to drive the *digital transformation* of the *public and private sectors*.

European Commission

DIGITAL WP 21-22 Data Spaces at a glance

Green Deal	Smart Communities	Mobility
Manufacturing	Agriculture	Cultural Heritage
Health, Genomics	Health, Cancer Images	Media
Financial	Skills	Language
Public Procurement	Security and Law Enforcement	Tourism

9

European Commission
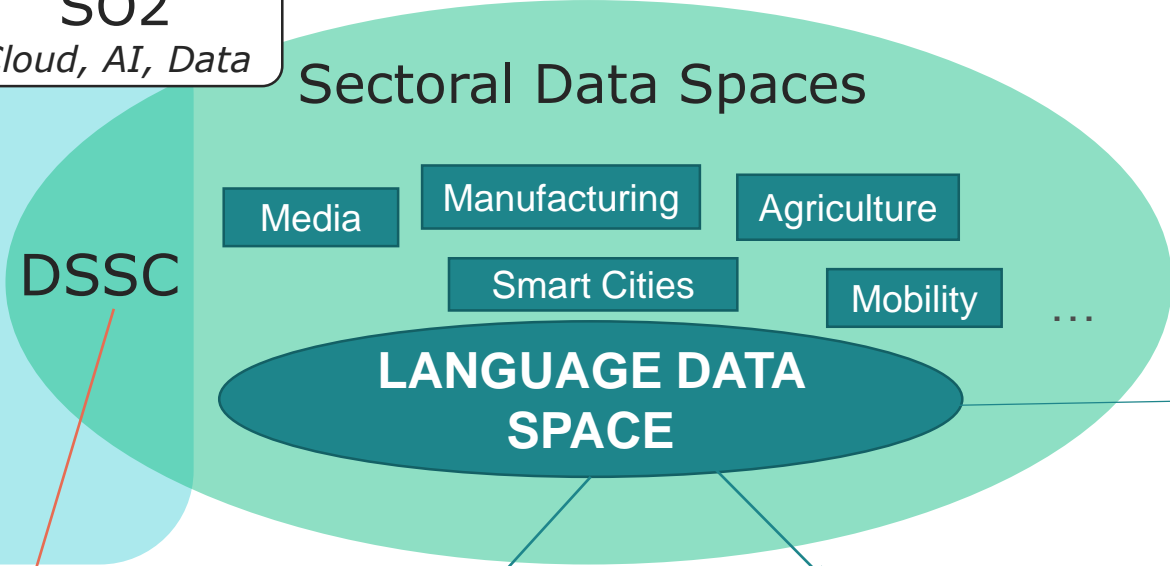
European DS Ecosystem – DIGITAL SO2

EDBI - European Data Innovation Board
supporting the EC in issuing guidelines on how to facilitate the development of DSs as well as identifying the relevant standards and interoperability requirements for cross-sector data sharing.

SO1
HPC

SO2
Cloud, AI, Data

SO4
Advanced Skills

European Data Spaces Technical Framework

consisting of

Cloud-to-Edge Infrastructures and Services

including

- Smart Middleware Solutions
- Marketplace

DSSC

Sectoral Data Spaces

Media

Manufacturing

Agriculture

Smart Cities

Mobility

…

LANGUAGE DATA SPACE

SO3
Cyber-security

DSSC – Data Spaces Support Centre
support all relevant actions on sectoral DSs and their architectures and data infrastructure requirements available for the DSs.

AI on-demand platform
connecting providers and consumers/users of AI resources.

TEFs – Testing and Experimentation Facilities
for experimenting with AI-based solutions in real-world environments.

SO5
Best use of Tech
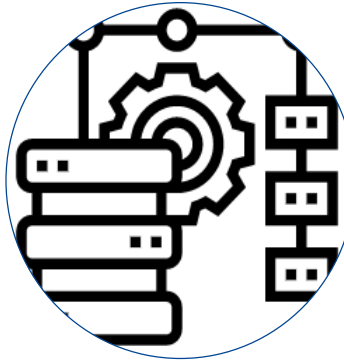**Common Services Platform**

Design Principles of the LDS

Data control

For data holders:
- Full control over data at all times;
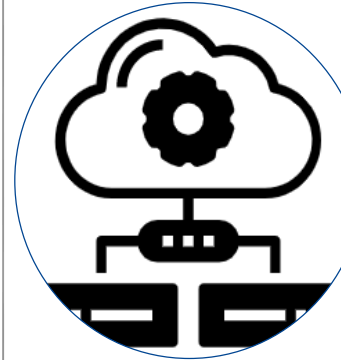- Option to change data access rights over time.

Governance Framework

- Fair,
- transparent,
- proportionate,
- non-discriminatory access to & sharing/use of data.

Respect of EU Rules & Values

- Personal data protection
- Consumer protection legislation
- Competition law

Technical Data Infrastructure

- Building blocks
- Efficient & coordinated DSs construction
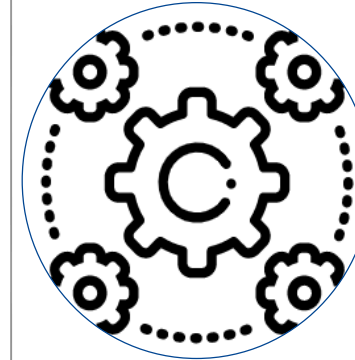
Interconnection & Interoperability

- Progressive interlinking of DSs
Aims:
- Fully integrated ecosystem
- Cross-exploitation of data

Openness

- For all actors, organisations or individuals;
Aims:
- Fostering competition
- Avoiding lock-ins
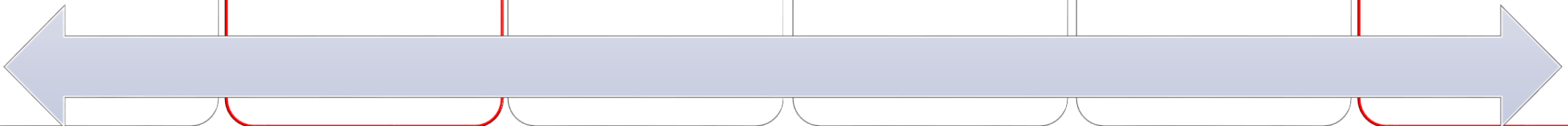
Financing

In DIGITAL WP **21-22**

Type of Action	PROCUREMENT
Indicative Budget	EUR 6 million
Indicative Time of the Call Opening	June – September 2022
Indicative Starting Date	Early 2023

In DIGITAL WP **23-24**

Planned types of actions: *procurement* & *grants*.

LDS Work Strands

Actions:

1 **Establishing** an institutional Centre of Excellence for Language Technologies (**CELT**)

Governance layer

→ to framework the collection, creation, sharing and re-use of **language data & models**, incl. legal aspects.

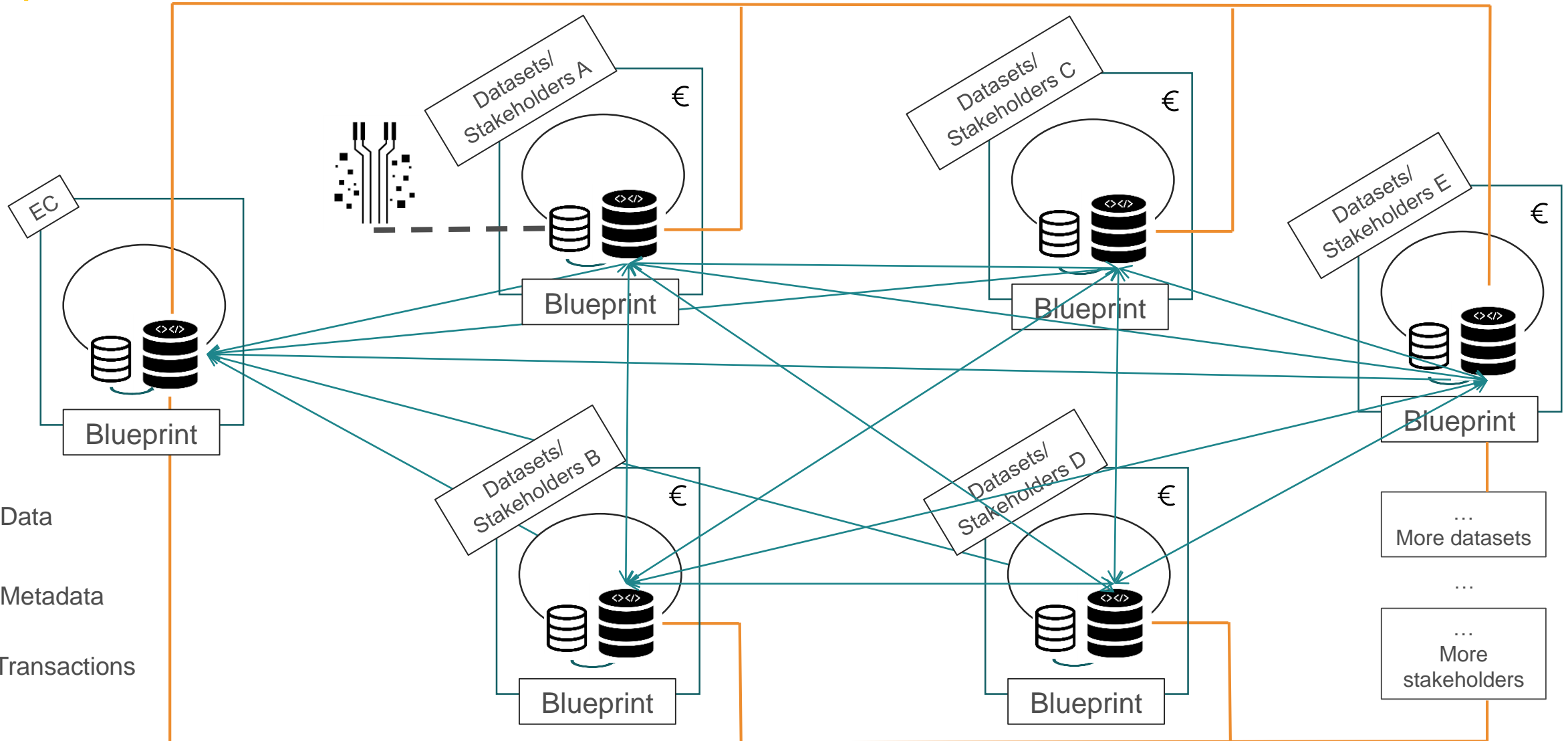**2** **Deploying** the Language Data Space

Operational layer

→ to implement the **architecture**;
→ to promote the **exchange of data**;
→ to support the deployment of **large multimodal language models** and a wide range of **AI language technology services** through the AI on-demand platform.

Establishment of the CELT & the CELT+

Strategic Level	**CELT** Coordinating across the MSs the creation and collection of multimodal language data and models.	*Networking*	Identifying and bringing together existing European **stakeholders/initiatives** and **language resources** • Datasets and data streams • *Large language models* *ELRC, EURAMIS, SCIC repositories, IATE, CLARIN, META-SHARE, etc.*
		Governance	Developing a **Multi-Stakeholder Data and Services Governance Scheme**
Operational level	**CELT+** Driven by Industry	*Planning*	Elaborating a **Blueprint for the Language Data Ecosystem** based on existing EU/national legislation, incl. a detailed **roadmap** on how to deploy the language data space

EC Role: Facilitator

European Commission

LDS Distributed Architecture

Datasets/ Stakeholders A

Datasets/ Stakeholders C

Datasets/ Stakeholders E

Datasets/ Stakeholders B

Datasets/ Stakeholders D

EC

Blueprint

€

...
More datasets

...

...
More stakeholders

Data

Metadata

€ Transactions

Stakeholders' Landscape

Other Data Spaces

Smart Middleware = cloud-to-edge infrastructure

Governance of the distributed data Infrastructure

PRIVATE

PUBLIC

Industry
- Text Audio Video
- (Language) **Models**

Media
- Text Audio Video
- **(Media) Streams**

Research Institutes
- Text Audio Video
- (Language) Models

Cultural Associations
- Text Audio Video

Public Administrations
- Text Audio Video

ELRC, EURAMIS, SCIC repositories, IATE, CLARIN, META-SHARE

AI on-Demand Platform

HPC, TEFs, EDIHs, …

Business Intelligence
Access to internal and international market(s)
Customised services
Economic Recovery — **Industry**

Business Intelligence
Access to networks of stakeholders and potential customers
Personalised services — Media

Access to data — **Researchers**

Better Information
Transparency
Access to digital services supported by language technologies — **Citizens**

More and better services offered
Digitisation — **Public Administrations**

16

European Commission
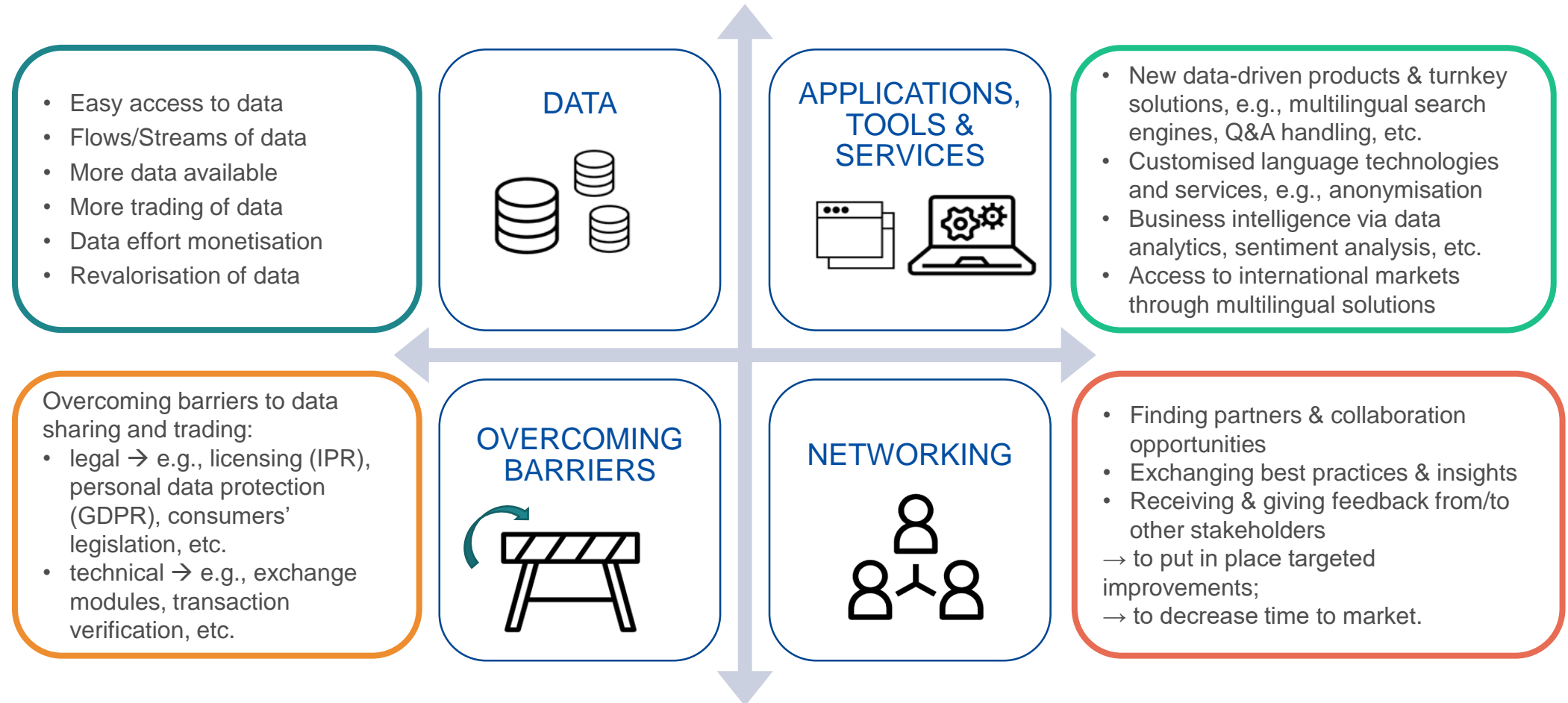
Benefits of the LDS

The LDS has been thought to be **a platform and a marketplace** *where the request for language resources and services matches the offer.*

- Easy access to data
- Flows/Streams of data
- More data available
- More trading of data
- Data effort monetisation
- Revalorisation of data

DATA

APPLICATIONS, TOOLS & SERVICES

- New data-driven products & turnkey solutions, e.g., multilingual search engines, Q&A handling, etc.
- Customised language technologies and services, e.g., anonymisation
- Business intelligence via data analytics, sentiment analysis, etc.
- Access to international markets through multilingual solutions

Overcoming barriers to data sharing and trading:
- legal → e.g., licensing (IPR), personal data protection (GDPR), consumers' legislation, etc.
- technical → e.g., exchange modules, transaction verification, etc.

OVERCOMING BARRIERS

NETWORKING

- Finding partners & collaboration opportunities
- Exchanging best practices & insights
- Receiving & giving feedback from/to other stakeholders
→ to put in place targeted improvements;
→ to decrease time to market.

# Any questions so far?

European Commission

Tour de table/Exchanges/Q&A

If you have sent us a PPT, **we will share** your slides on screen. If not, on the next slide, there are a couple of **questions** to give your contribution.

Speakers' names will be displayed on screen in **random order**.

Please **speak up** when you see **your name.**

If you have any **questions**, please **feel free to ask** them after each presentation.

European Commission
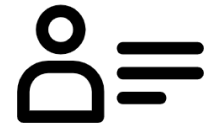
Your Business & Data, Models, Tools, Services & the LDS

Some questions

- What **kind/s** and **amount/s** of **data** do you handle?
- How are they handled, e.g., in terms of **privacy** and **security**?
- Do you use any **tools** or **services** based on **language technologies**?
- Do you use **language models**? If so, what types?
- What are the **main difficulties** concerning tools, services and models?

- **Pros and cons** of the Language Data Space so far?

- Any **major challenges/risks** connected to the Language Data Space and possible solutions?
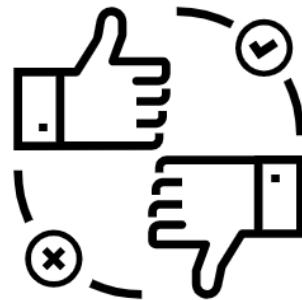
Please feel free to ask your own questions!

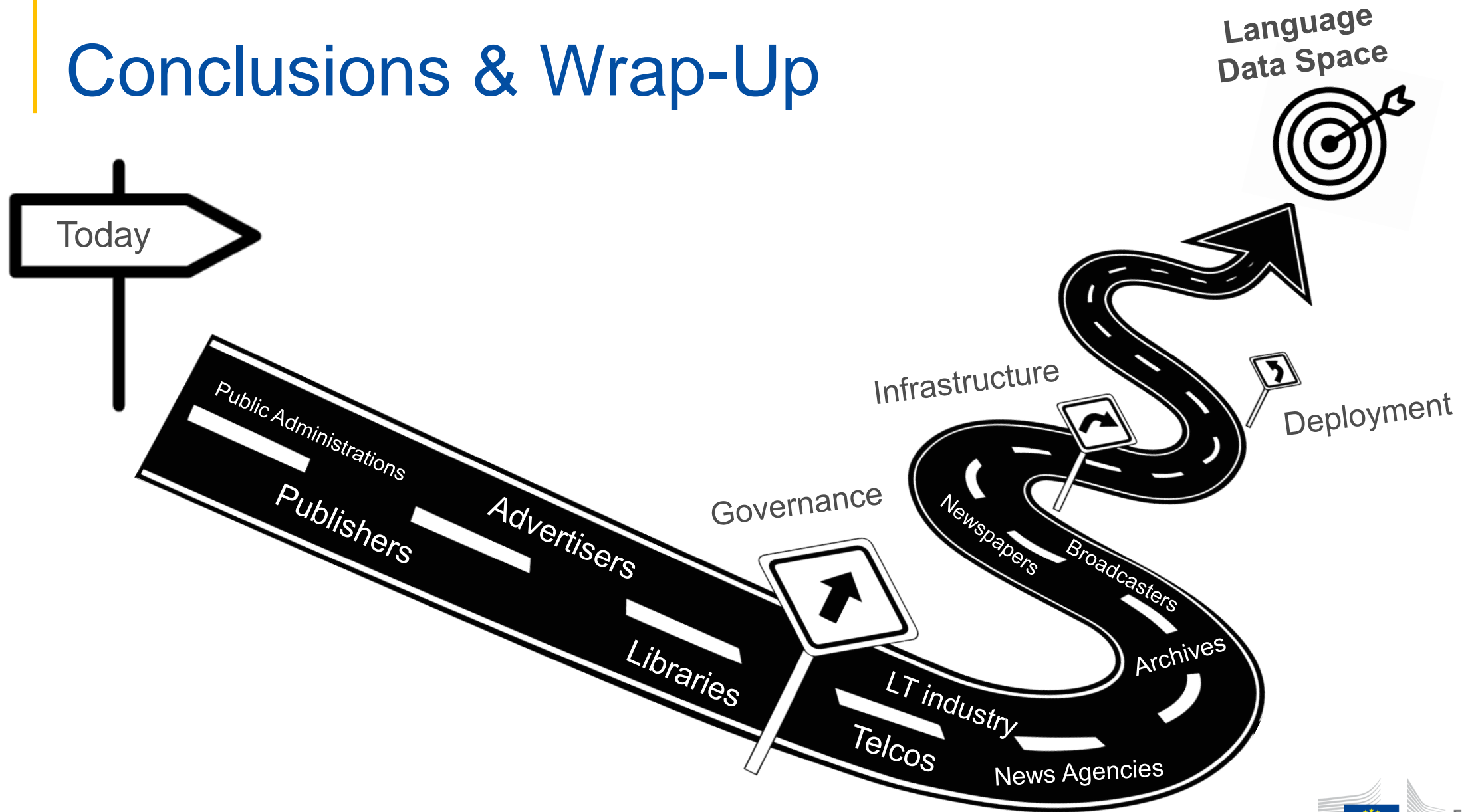European Commission

Any other aspects you would like to mention or any other questions?

European Commission

Conclusions & Wrap-Up

Today

Language Data Space

Public Administrations

Publishers

Advertisers

Libraries

Governance

Infrastructure

Deployment

Newspapers
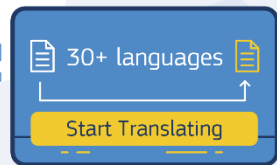
Broadcasters

Archives

LT industry

Telcos

News Agencies

European Commission

Thank you very much for sharing your ideas!

For any follow-up questions, remarks and/or queries, please feel free to send us an email at: CNECT-LANGTECH@EC.EUROPA.EU
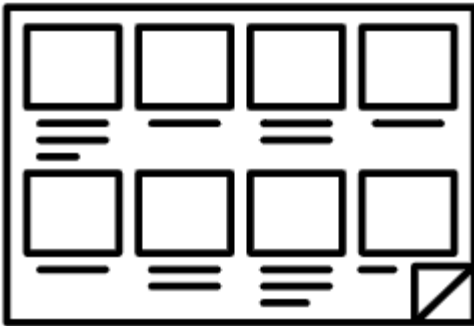
Need eTranslation or other language tools?
Visit: https://language-tools.ec.europa.eu/

30+ languages

Start Translating
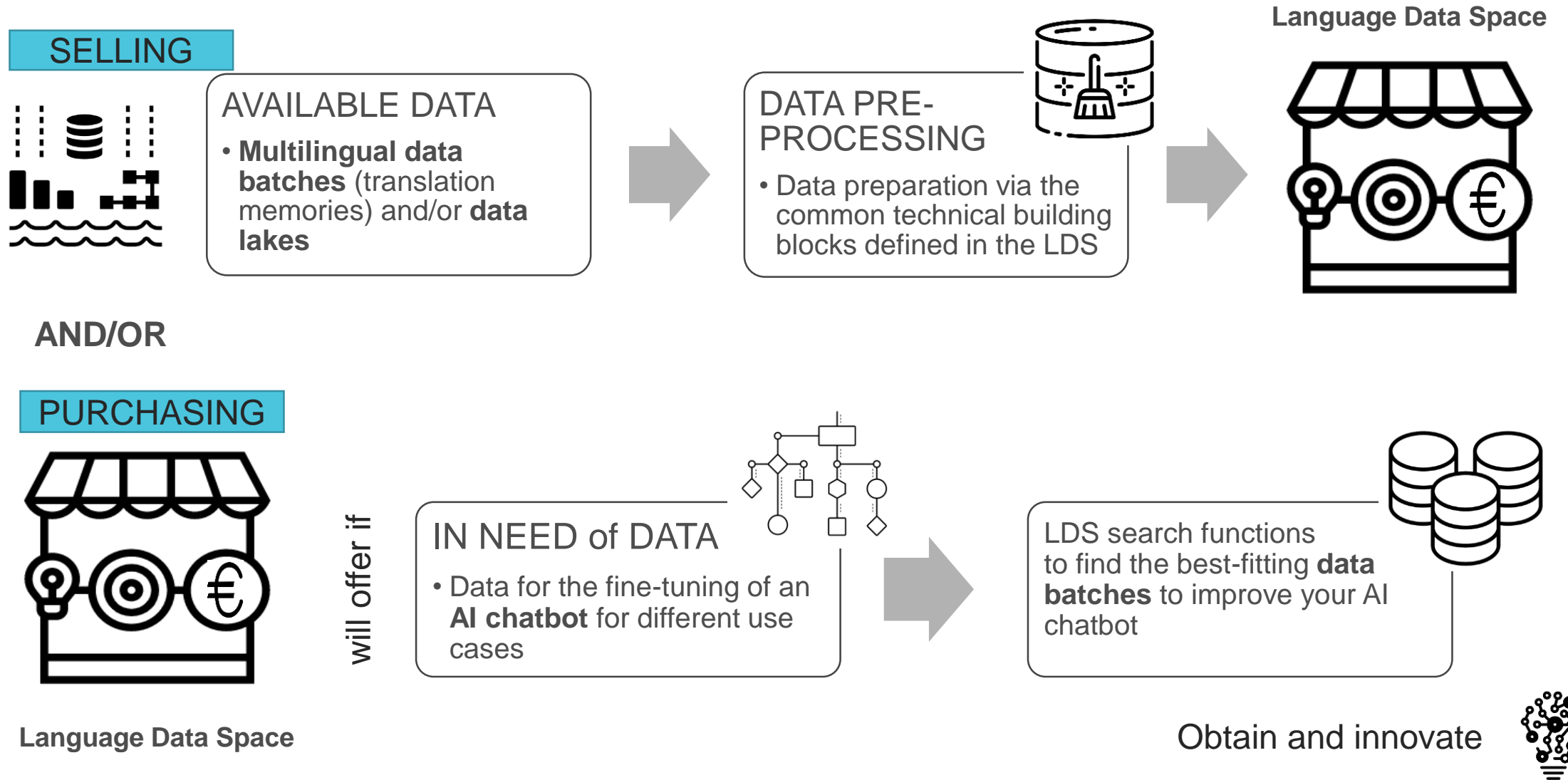
eTranslation

European Commission

Optional Slides

Examples of Use Cases

- Scenario 1 – Data Sharing
- Scenario 2 – Applications, Tools and/or Models
- Scenario 3 – Services
- Scenario 4 – Data Analytics
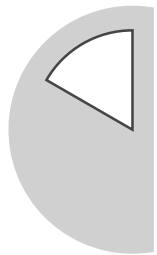- Scenario 5 – Market Exploitation

European Commission

Specific Benefits: Scenario 1 – Data Sharing

SELLING

AVAILABLE DATA
- **Multilingual data batches** (translation memories) and/or **data lakes**

DATA PRE-PROCESSING
- Data preparation via the common technical building blocks defined in the LDS

Language Data Space

AND/OR

PURCHASING

will offer if

IN NEED of DATA
- Data for the fine-tuning of an **AI chatbot** for different use cases

LDS search functions to find the best-fitting **data batches** to improve your AI chatbot

Language Data Space

Obtain and innovate

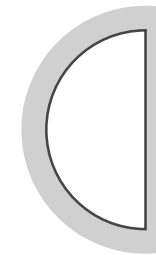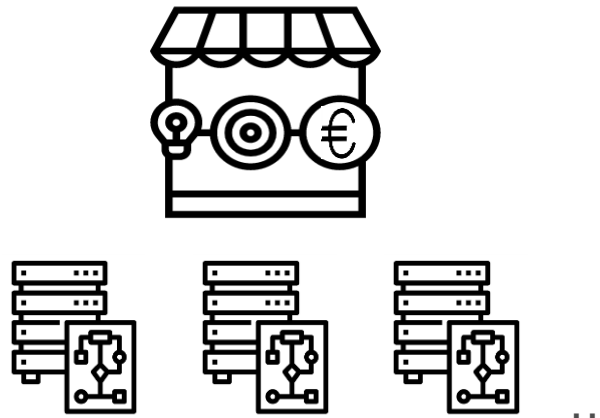# Specific Benefits: Scenario 2 – Applications, Tools and/or Models
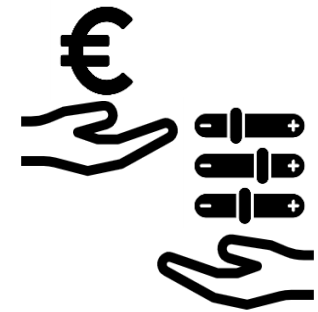
Automatic Speech Transcription

You want to build your own automatic speech transcription workflow to cover different market use cases.

You can go to the LDS and look for algorithms or pre-trained models capable of converting speech or audio into a written format.

You can then improve them and make them available in the LDS marketplace for further business.

...

Specific Benefits: Scenario 3 – Services

Automatic Summarisation as a Service

You can use the LDS to **build or better** any of the following aspects in the automatic summarisation pipeline to offer **a higher-performing service** to new or existing customers:
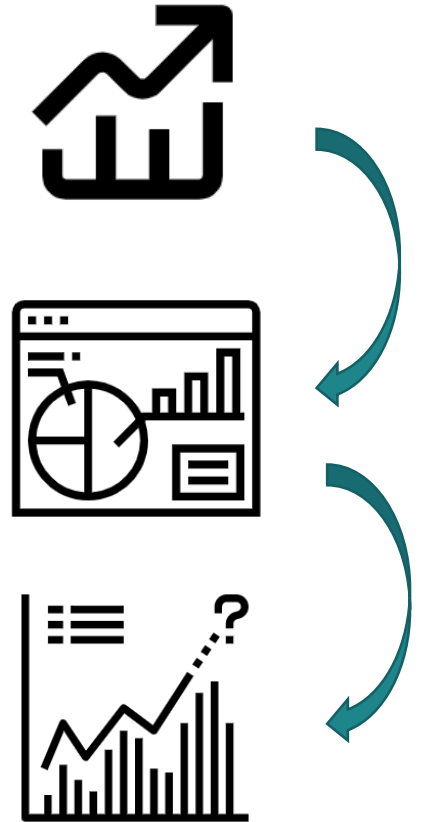
- Single-document versus multi-document summarisation
- Extracts versus abstracts
- Generic versus user-focused summaries
- Indicative versus informative summaries
- Language of summaries
- …

European Commission

Specific Benefits: Scenario 4 – Data Analytics

You have historical and transactional data and need to run predictive analytics on it.

You can go to the LDS and look for a statistical/visualisation library.

You can gain insights from your data and make predictions.

European Commission

Specific Benefits: Scenario 5 – Market Exploitation

LT / ZH — You have new stakeholders from **Lithuania** and **China**.

You **train** your MT engines.

You **re-sell** these trained systems to create further value for your effort.

You want to add Lithuanian and Chinese to your **MT engines** to have certain elements/content automatically translated on your website.

You purchase the data you want or obtain it for free in case of open data.

Your website will be **open** to **more users** in the EU and on the global market.

You can go to the **LDS**.

You will find **thesauri, terminology databases, multilingual parallel corpora**, etc. to train your MT systems.

European Commission